7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

1

Creating the Dataset for the Western Wind and Solar Integration Study (U.S.A.)

Cameron W. Potter, Debra Lew, Jim McCaa, Sam Cheng, Scott Eichelberger and Eric Grimit

***Abstract*--The Western Wind and Solar Integration Study (WWSIS) is one of the world's largest regional integration studies to date. This paper discusses the creation of the wind dataset that will be the basis for assessing the operating impacts and mitigation options due to the variability and uncertainty of wind power on the utility grids. The dataset is based on output from a mesoscale numerical weather prediction (NWP) model, covering over 4 million square kilometers with a spatial resolution of approximately two-kilometers over a period of three years with a temporal resolution of 10 minutes. The mesoscale model dataset includes all the meteorological variables necessary to calculate wind energy production. Individual time series were produced for over 30 thousand locations representing more than 900 GW of potential wind energy generation.**

Index Terms— data processing, meteorology, power system meteorological factors, power system modelling, power system planning, weather forecasting, wind energy

I. INTRODUCTION

WEATHER-DRIVEN renewable energy sources require a new paradigm in power systems analysis. Conventional fossil fuel power plants can be operated in accordance with the needs of the power system. Renewable energy sources such as wind or solar are variable and thus the operating schedules of such plants are largely dictated by the changing "fuel" supply. This is especially pertinent in the case of wind, photovoltaic solar and run-of-the-river hydro, none of which have inherent storage in their power plant design. This variability may result in increased costs, largely manifested through an increase in the ancillary services and/or regulation reserve required to maintain power system reliability. Integration studies assess these operating impacts and their associated costs and require a solid understanding of the varying fuel supply.

Various methodologies exist for conducting grid integration studies for wind power and these are getting significant attention. However, the importance of the fuel supply data that are used in these studies is often overlooked. The Utility Wind Interest Group released a report in 2006 [1] on the state-of-the-art in wind integration that identified many key considerations in trying to integrate wind into a power system. However, the report did not mention the need for greater accuracy in the simulation of synthetic power output. The International Energy Agency (IEA) released a report on the research and development needs for wind energy [2] that identified the need for the creation of a wind characteristics database, but this database primarily focuses on wear and tear on wind turbines rather than their characteristic power generation [3]. Follow-up work by the IEA with the Organisation for Economic Co-operation and Development (OECD) [4] recognised the need for wind energy forecasting, but still overlooked the need for accurate modeling of synthetic wind energy. The European Wind Integration Study [5] used observed data as the basis for its prospective build-out scenarios. This may be reasonable for limited scope scenarios that primarily assess the expansion of existing wind projects, but is not sufficient to accurately assess the effects of geographical diversity when new wind projects are studied.

As integration studies become more sophisticated, the importance of modeling the supply is increasingly recognized [6]. Ideally, an integration study would be based on observed power output data for each project, yet since these studies are used for planning purposes, they must rely on model output for projects that are not yet built [7]. In fact, even long-term, on-site meteorological data (as opposed to power data) are rarely available. Therefore, a different technique must be used to synthesise the behaviour of renewable energy projects to be considered in the integration study. IEEE Transactions on Power Systems had a Special Section on Wind Energy in 2007 including the paper "Utility Wind Integration and Operating Impact State of the Art" [7] which stated:

> "*A state-of-the-art wind-integration study typically devotes a significant effort to obtaining wind data that are derived from large-scale meteorological modeling that can re-create the weather corresponding to the year(s) of load data used.*"

The use of meteorological models and related post-processing in support of wind integration studies is the topic of the current paper, which uses the Western Wind and Solar Integration Study (WWSIS) as a case study. The fundamental goal of the WWSIS, which was funded by the Department of Energy (DOE) and coordinated by the National Renewable Energy Laboratory (NREL), was to produce a comprehensive dataset that could be used to model the build-out of potential wind plants in the western United States. This study is on of the largest wind integration studies to-date, covering an area of more than 4 million square kilometers modeled at a two-kilometer

---
C. W. Potter is with 3TIER, Seattle, Washington, USA, (e-mail: cameronwpotter@ieee.org).
D. Lew is with the National Renewable Energy Laboratory, National Wind Technology Center, Boulder, Colorado, USA, (e-mail: debra_lew@nrel.gov).
J. McCaa is with 3TIER, Seattle, Washington, USA, (e-mail: jmccaa@3tiergroup.com).
S. Cheng is with 3TIER, Seattle, Washington, USA, (e-mail: scheng@3tiergroup.com).
S. Eichelberger is with 3TIER, Seattle, Washington, USA, (email: seich@3tiergroup.com).
E. Grimit is with 3TIER, Seattle, Washington, USA, (e-mail: egrimit@3tiergroup.com).

7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

2

resolution, resulting in over 1.2 million grid points. From this base dataset, over 30,000 points were selected for further evaluation. Each of these points was modeled as an individual project with ten VestasV-90 (3MW) turbines. The cumulative amount of wind energy modeled for the entire study is over 900 GW. This dataset will be made public via a web-interface hosted by NREL and designed by 3TIER.

II. METHODS FOR WIND DATA CREATION

The lack of high quality wind resource data that are synchronized with the region's electric load is one of the primary obstacles to conducting wind integration studies. Time series data must be used to perform power system analysis for systems with significant wind penetration [8]. These data can be obtained in one of three ways: 1) on-site observations; 2) data mining of either offsite observations or reanalysis data; and 3) mesoscale modeling [6].

On-site observations of power generation are the most desirable data source because they represent what *actually* happened (within the accuracy of the instrumentation/data storage). However, power output observations can only be obtained from existing wind projects – meaning that the integration studies would have limited application. Alternatively, wind speed observations could be used, but these are not quite as accurate as observations of power output, because the conversion of wind speed to power output is not entirely deterministic, even at the turbine level. Furthermore, a wind speed measurement at a single point does not adequately represent the average wind speed across the entire project. The project average wind speed and power output tend to be smoother than the same quantities at a single location or turbine [7]. Multiple hub-height anemometers during the entire period of interest are required to obtain good information for a single site and although this information exists for some locations, it is not available for most sites. Thus, the use of on-site observations is of limited use for integration studies.

Data mining/data manipulation is another method of obtaining data for an integration study. Measure, correlate, predict (MCP) is the most prevalent data mining method used to produce synthetic wind or wind power time series data. MCP takes a short-term record of on-site measurements and correlates it to a long-term record of measurements at an off-site observation station. This technique has some serious flaws that render it inappropriate for most integration studies. First of all, it still requires (short-term) on-site observations to establish the correlation to the off-site data. This prevents an analysis of sites for which on-site data are not readily available. A second major limitation is that a nearby off-site observation tower with a sufficiently long measurement period must exist. Although a sufficient number of long-term observation stations exist in some areas, this is generally not the case in sparsely populated regions, or in most developing countries. A third limitation is that off-site observations generally have a temporal resolution of one hour, but that higher temporal resolution data are desirable for most integration studies. In addition, the accuracy of MCP can be problematic if the on- and off-site locations do not have similar meteorological characteristics. Most long-term weather observation towers are located in places such as airports and airports are intentionally built in low wind speed locations. The local weather phenomena can be markedly different for sites that are relatively close, especially if one site is specifically chosen for low wind speeds and the other site is chosen for high wind speeds. Complex terrain further amplifies this source of error. However, even if all of the data are available and the MCP technique is appropriate for the site of interest, it still suffers from the limitation that MCP cannot be used to produce a gridded dataset. It may be able to produce a synthetic time series for a single location, but time series at multiple locations are needed to model the effect of smoothing across a large wind project [9].

Data mining of reanalysis data [10] (spatially and temporally coarse global datasets) avoids the need for nearby, long-term, off-site observations, but does not provide sufficient temporal resolution and still suffers from many of the same flaws as MCP – such as unreliable correlations and the lack of ability to model multiple points per farm (with non-trivial smoothing effects).

Numerical weather prediction (NWP) models are a good alternative to data mining or on-site observations. The NWP simulations are driven by conservation equations that model the physical interactions in the atmosphere. The NWP models employ the reanalysis wind speed datasets (mentioned previously) to determine boundary conditions for the model run, which is then realistically downscaled (using physical equations) to a finer physical resolution. With sufficient computing power, these models can be used to calculate wind speeds at evenly spaced grid points over a very wide area and can also produce simulations at several heights above the surface. In addition, since the models are reproducing the physical interactions in the atmosphere, there are no inherent temporal limitations. Wind speed datasets with a temporal resolution of ten minutes can be obtained with the boundary conditions provided by a long-term reanalysis dataset with a temporal resolution of six hours.

The NWP model that downscales the reanalysis data is termed a mesoscale model. Because the mesoscale models run over a smaller area than the larger synoptic scale models, which are often employed by the weather bureaus, the physics of the model can include additional detail. For instance, the larger models assume hydrostatic conditions ignore the effects of local topography and land use. In contrast, mesoscale models can be non-hydrostatic and can simulate smaller scale wind patterns such as thermally driven local winds and the Venturi effect (mountain winds). In fact, the output from mesoscale models can be dominated by terrain and non-hydrostatic phenomena. These winds may not be accurately modeled through data mining from off-site observations or data mining based on reanalysis data alone.

Once the weather data have been downscaled to a finer temporal and spatial grid, the key parameters (such as wind speed) can be compared with shorter-term observations, possibly at multiple locations. Model output and observations often differ due to the necessary simplifications that the model makes in representing the complexity of the atmosphere. These errors can be reduced with Model Output Statistics (MOS) equations. The MOS equations are used to make statistical adjustments to a

7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

3

modeled dataset. These corrections are possible because the gridded data cover an area in a continuous manner and observation data at one site will have a sphere of influence that can be used to adjust the model outputs.

However, the NWP methodology is still a model and has two drawbacks. Firstly, the NWP mesoscale models do not accurately represent all weather patterns and some regions can have significant errors, which cannot always be corrected with MOS equations. However, experience shows that these locations are the exception rather than the rule. The second drawback with mesoscale modeling is the large computational requirement. Large area NWP simulations with a sufficiently high spatial and temporal resolution require large supercomputing facilities and even then are slow to operate. Nevertheless, the flexibility and accuracy of mesoscale modeling justifies the use of the methods and computing resources for integration studies. Even more so since the alternatives cannot provide the gridded data required for a large wind integration study.

III. WWSIS Dataset Generation

The Western Wind dataset was created in two separate stages, but the modeling technique was consistent to allow for a smooth combination of the datasets. The first stage modeled the Pacific Northwest and was performed for the Northwest Wind Integration Action Plan (NWIAP) [11], jointly sponsored by the Bonneville Power Administration (BPA) and NREL. It covered the states of Washington, Oregon and Idaho as well as most of Montana and Wyoming. Fig. 1 shows the area covered by the NWIAP modeling effort bounded by a striped box. The second stage expanded the modeling area to include most of the western United States west of 100°W longitude.
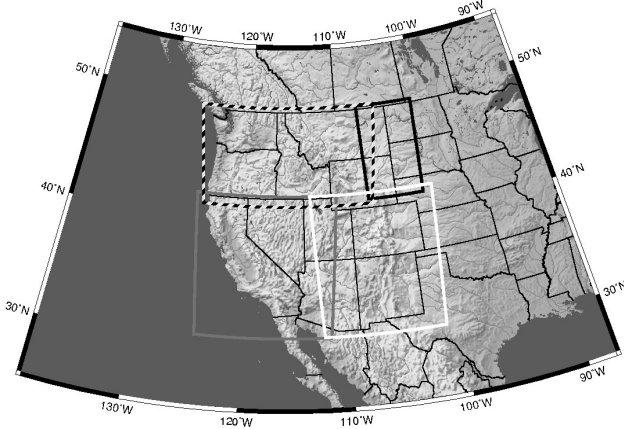

Fig. 1. A map showing the modeling domains in the WWSIS. The striped bounding box shows the NWIAP region and the other domains, black, grey and white, are called Domains 1, 2 and 3 respectively.

A. Model Domains

Fig. 1 shows four domains: the NWIAP domain, and three other domains. The use of multiple domains was forced by the magnitude of the area that was modeled at a high resolution.

The model runs are often too large (especially in this case) to run in the memory of a single processor, The simulation is parallelized by allocating sub-sections of each of the model domains (sub-domains) to individual computer processors on a supercomputing cluster. However, the processors that simulate each of the sub-domains cannot do

the calculations entirely independently. Each processor must communicate with the other processors for adjacent sub-domains. This is required to allow "advection" and "diffusion" operators to transfer information about weather events between neighboring sub-domains.

Sub-domains allow these models to run accurately and relatively quickly, but in practice the number of sub-domains that can be accommodated is still limited. The size of each sub-domain is memory-limited and the number of sub-domains is limited by the bandwidth of the inter-node links on the compute cluster. If too many sub-domains are used, the communication channels in the cluster become clogged, resulting in increased latency. The southern region identified in Fig. 1 had to be split into two domains to prevent potential latency problems with the compute cluster.

B. Model Configuration

The Weather Research and Forecasting (WRF) model [12] is generally considered to be the most advanced mesoscale model in North America and has superseded the previous industry standard, the MM5 model [13]. Thus, the WRF model was used to perform the mesoscale modeling for this project. The WRF model can be configured to better represent the physical processes based on model domain, resolution and application. Four different model configurations were tested for the WWSIS model simulation, the configurations are shown in Table I.

TABLE I
NWP Configurations Using the Advanced Research WRF Core

	Vertical Levels	Planetary Boundary Layer Parameterisation	Elevation Dataset	Land Surface
A	31	Yonsei University	30 arc-second USGS	5-layer soil diffusivity
B	31	Mellor-Yamada-Janjic	30 arc-second USGS	5-layer soil diffusivity
C	31	Yonsei University	30 arc-second USGS	Oregon State Uni.
D	37	Yonsei University	30 arc-second USGS	5-layer soil diffusivity

Configuration A was used as the baseline model configuration with configurations B, C and D all having a single parameter of deviation. Configuration B used the Mellor-Yamada-Janjic boundary layer parameterisation, which features explicit prognostic equations for boundary layer turbulence. Configuration C used the Oregon State University land surface model, a more sophisticated physical process model for estimating surface fluxes. Both Configurations B and C should theoretically be better than Configuration A. However, the increased sophistication in the models introduces additional assumptions and unconstrained parameters that can adversely affect the accuracy of the model. Configuration D adds extra vertical levels in the boundary layer to better simulate the vertical profiles of wind and temperature near the surface.

The trial runs that evaluated the various configurations were simplified, because of the computational cost of the simulations. The trials were run at a coarser spatial resolution of 6 km × 6 km grid spacing instead of 2 km × 2 km and the model was only run for three out of every nine days for the year 2006.

The four different model configurations were run for each of the domains. The NWIAP domain was modeled first and validated against six tall towers. The validation showed

7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

4

that the default configuration, A, was optimal. The other three domains were validated against a total of 30 tall towers. Each of the different configurations was judged qualitatively "best" (over a number of parameters) for at least one tower. For Domains 2 and 3 Configuration D outperformed the other configurations most consistently. For Domain 1 Configurations A and D performed at a similar level of accuracy. The study team decided to use Configuration D for Domains 1, 2 and 3 to assure consistency.

C. Dataset Creation

The selected models were run on the supercomputing cluster. For each model grid point, defined by latitude, longitude and elevation, the following twenty-one ten-minute time series were archived:

- wind speed and direction at 10 m, 20 m, 50 m, 100 m, 200 m above the surface as well as at 500 hPa (higher in the atmosphere);
- temperature at 0 m, 2 m, 20 m and 50 m above the surface;
- specific humidity at 2 m above the surface;
- pressure at the surface;
- precipitation at the surface;
- downwelling radiation (longwave and shortwave) at the surface.

D. Dataset Regridding

The model run was performed at 2 km × 2 km grid spacing across each domain. However, the edges of the domains were not perfectly aligned as each domain was defined individually. The original datasets were regridded to achieve a consistent grid spacing across the entire area covered by the WWSIS. The final dataset has a one arc-minute spacing, which allows easy identification of the grid points by latitude and longitude.

E. Dataset Blending

The four single domain simulations needed to be blended to produce a single, consistent, dataset for 2004-2006 with a temporal resolution of ten minutes and a spatial resolution of one arc-minute. In order to produce a seamless dataset, data from the individual model domains were blended at the overlapping boundaries (see Fig. 1.). The result was a single large dataset with over 1.2 million individual grid points. Each of these grid points had an associated time series with 157,680 time steps for each of the parameters listed in *Section IIIC. Dataset Creation*. This dataset, stored in netCDF format, used more than 24TB of storage space.

The sheer size of this dataset caused significant problems. To maintain the integrity of the dataset, the dataset was copied each time a process was implemented that altered the core dataset (e.g. re-gridding, blending, etc.). The copy was then altered and the original dataset was maintained until the altered duplicate could be thoroughly verified. This meant that for much of the time many TB of duplicate data were being stored as a safety backup. This process was difficult and time consuming, as even the process of copying 24TB of data is non-trivial. However, the production of the dataset was a major cost of the project (both in time and money) and loss of the dataset was not an acceptable risk.

IV. POST-PROCESSING THE WWSIS DATASET

The creation of the modeled dataset was the first phase of the project, however, the modeled dataset had to be converted into synthetic wind energy project data to make the data easily accessible for power systems modeling. For the purpose of creating these synthetic wind projects, each grid point was assumed to be its own potential wind project.

A. Site Selection for Synthetic Wind Energy Projects

Ideally each grid point in the modeled dataset would be converted into a synthetic wind energy project. However, many locations are not suitable for the location of wind energy projects due to other uses, building restrictions and an inadequate wind resource. In addition, such an approach would have been impractical given the large scope of this study. The computational time required to access the data, convert from wind speed to effective wind speed (adjusted by air density) to wind power for each of the 1.2 million points at almost 160,000 time intervals would have been large. Furthermore, even if each grid point were only assumed to support a single utility-scale turbine the result would still be several TW of wind energy, well beyond the scope of likely wind energy build-out scenarios. Thus, a subset of the potential sites was selected for modeling as synthetic wind projects.

The power systems portion of this integration study requires approximately 70 GW of installed wind energy. However, to allow evaluation of a large number of different build-out scenarios, it was decided that 3TIER would provide time series data for over 900 GW of synthetic wind energy sites. To determine the number of MWs that each site could represent some simple heuristics were used: a spacing of ten rotor diameters between strings of wind turbines and a spacing of three rotor diameters between turbines on the same string.

To maintain consistency across the dataset, the same turbine was used for each synthetic wind project. A large turbine was employed because the dataset was designed to represent build-outs of wind energy up to 2017 (ten years in the future from the commencement of the project) and there is a trend towards larger turbines. The Vestas V-90 3MW turbine was chosen as a good middle ground between today's mean turbine size and those likely to be used in the future. Using the simple heuristics described above, ten turbines could be assigned to each grid point representing a total of 30 MW of installed capacity for each site. As a result, 30,000 points were required to model the total amount of 900 GW. Multiple sites could then be aggregated to obtain wind energy projects of a larger size, still modeled to allow for varying wind speed across the project.

The site selection process was carried out in several phases. During each phase additional potential project locations were added to meet a specified goal. The main purpose of the WWSIS was to model the WestConnect group of utilities (excluding California). These utilities are in Nevada, Arizona, New Mexico, Colorado and Wyoming. However, the entire Western Electricity Coordinating Council (WECC) area was modelled to allow for interactions at the borders of the WestConnect footprint. The first phase was to pre-select a set of points to represent existing wind energy projects and those under development.

7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

5

This information was obtained and compiled by NREL and resulted in 404 sites (or approximately 12 GW).

The next phase was to identify the sites with the highest wind energy potential (based on wind energy density at 100 m) within 50 miles of existing or planned major transmission networks or in pre-identified high potential renewable energy zones (REZ) in the study footprint. 200 GW of sites (6667 sites) were selected in the transmission corridors or REZ areas.

The third selection phase aimed to find the sites that had the best correlation with the load profile of the West Connect (limited to sites with a wind energy density of greater than or equal to 300 W/m^2). The load correlation measure was evaluated by calculating the difference between the average normalised load profile and average normalised wind energy density (on an hourly basis) – the smaller the difference, the better the site. It was also desirable that the sites were geographically diverse; this was achieved by assigning each state (and two offshore regions) an approximate number of GW that should be selected. These assignations were based upon the "20% Wind Energy by 2030" analysis performed by NREL and others [14] and also the relative importance of each state to the WWSIS study (primarily focusing on the WestConnect region). Table II shows the approximate GW modeled in each state.

TABLE II
DISTRIBUTION OF SITES FOR SELECTION (BY STATE AND OFFSHORE REGION) USING LOAD CORRELATION AND POWER DENSITY

State/Offshore Region	Selected by load correlation [GWs]	Selected by power density [GWs]
Arizona*	18	18
California	8	74
Colorado*	28	28.5
Idaho	8	13.5
Montana	13	35
North Dakota	4	5
Nebraska	8	5
New Mexico*	32	40.5
Nevada*	33	48
Oklahoma	7	7
Oregon	4	36
South Dakota	7	10
Texas	8	10
Utah	8	11
Washington	4	44
Wyoming*	54	69
Offshore CA	1	4
Offshore WA/OR	0.5	1
TOTAL	245.5	459.5

*In the West Connect study footprint

The fourth selection phase was a simple selection by highest wind energy density, again selected according to the allocations in Table II. The selected sites are shown in Fig. 2. Finally, after the site selection was complete, it became apparent that some sites that should have ideally been included in the pre-selected set of sites had been missed. A further set of "post-selected" sites was identified with input from project stakeholders resulting in an additional 1499 points. The final number of points that were selected for further study was 32043, each representing a single 30 MW generation site.
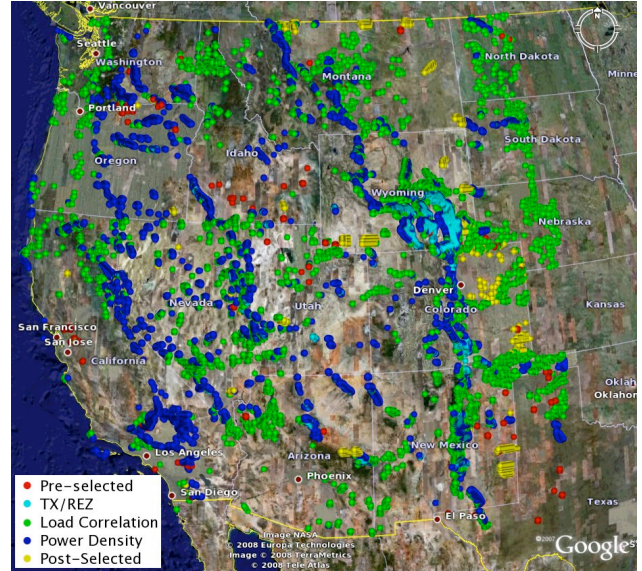

Fig. 2. A map showing the selected sites with each point coloured differently depending on selection technique.

B. Creation of a Wind Power Time Series at Each of the Selected Points

Numerical weather prediction models have a tendency to produce wind speed time series that are excessively smooth, that is, they do not produce sufficient wind speed variation at short timescales. As a result, wind plant output derived directly from wind speeds from a mesoscale model and put through a rating curve is excessively smooth. Unfortunately, this simple conversion technique is still regarded by most as the industry standard. An alternative technique, the Statistical Correction to Output from a Record Extension (SCORE), was proposed in a paper presented at the IEEE Power Engineering Society General Meeting in 2007 [6]. SCORE has now been used for five different studies, modeling several GWs of potential wind energy installations. The SCORE process uses observed statistical deviations from a mean value to create probability density functions of deviation from some central point. SCORE is run for each individual turbine and produces a time series of data for each turbine. The individual turbine time series are then aggregated to represent sub-project groupings or summed up to model the entire project output. However, use of a probabilistic process to model the output for 32043×10 individual turbines is extremely time consuming. In addition, the turbine locations within each 2 km × 2 km grid would only be approximate, meaning that the individual turbine locations would provide no extra information. To solve this problem SCORE-lite was developed.

SCORE-lite models each grid point, instead of each turbine, by aggregating ten individual samples from the original SCORE probability density functions (as though ten turbines were being modeled) to develop new probability density functions that represent ten turbines instead of one. The goal of SCORE-lite is to take the "rated" power output, calculated by converting wind speed to power output through a simple rating curve, and modify it such that the overall ramping characteristics more closely approximate those observed in reality. SCORE-lite was validated as part of this project and found to result in a more realistic number of ramps without any appreciable loss of accuracy in modeling the diurnal cycle.

7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

6

V. FORECASTS FOR THE WWSIS DATASET

A wind energy forecast was required at each synthetic wind energy site to adequately model operation of the power system with the hypothetical wind plants. Many studies have shown that accurate wind energy forecasts can reduce the costs of integrating wind energy into a power system [9, 15-19]. To adequately assess the costs and impacts of wind integration, the wind energy forecast plays a major role. Consequently, four forecasts were provided as part of the final dataset for this project. These four forecast methodologies represent the range of forecasting possibilities.

A. Persistence Forecast

A persistence forecast provides the simplest kind of forecast, but is only appropriate for short-term forecasting. As part of this study, the persistence forecast provided a one-hour forecast with a two-hour look-ahead period. This time delay was chosen as a representative delay as it allowed time for the forecast to be created and inspected, while still allowing time for an operator to react before the power had to be scheduled on the hourly timescale. For forecasts with a target period further in the future than the hour-ahead scale, other techniques must be used.

B. Climatological Forecast

A climatological forecast is also a very basic forecast. It is used for day-ahead prediction and is designed to capture the average hourly diurnal cycle for the present weather regime. The previous thirty-day period is often selected as the averaging period. For this project, each month of each year had its own climatological trace of 24 one-hour values. This actually includes "future" information in the forecast and so is not possible in reality. However, the climatological forecast is only a baseline forecast. The mesoscale model forecast provides more accurate forecasts than the climatological forecast.

C. Mesoscale Model Forecast

The mesoscale model forecast represents the state-of-the-art in day-ahead forecasting. This model forecast represents baseline accuracy for mesoscale model forecasting, as it is not tuned to any specific project. A mesoscale model forecast is run in a very similar method to *Section III – WWSIS Dataset Generation*.

The reason that a mesoscale model could be used to create the synthetic data as well as the forecasts is that different data were used to provide the boundary conditions to drive the mesoscale model. The NWP simulation for synthetic data creation was driven using the reanalysis dataset described above. The mesoscale modeling forecast was driven using a different input dataset, the Global Forecast System (GFS) [20], the actual information used to perform state-of-the-art forecasting.

The mesoscale model forecasting was meant to be a smaller portion of work than the simulation of synthetic wind energy data, so the same granularity of the models could not be afforded. Instead, the models were run with a 6 km × 6 km resolution and at the hourly timescale. This meant that the forecast model was coarser than the original model. As a result, less computationally expensive and could be run as a single large domain.

True state-of-the-art forecasting is specifically tuned to operate optimally at the desired forecast location through the use of a MOS correction. Due to the large number of sites (over 32,000 sites) such a detailed procedure was impractical. The mesoscale forecast is a good measure of forecasts obtained from a state-of-the-art model and also highlights characteristic errors – but it is not as good as a true state-of-the-art forecast.

D. "Perfect" Forecast

The "perfect" forecast is an artificial forecast that cannot be produced in reality, but can be used to find the minimum wind integration cost. Wind is a variable resource and so even if it is forecast perfectly, the resulting variation will still require some of the generators on the system to operate away from maximum efficiency (or change the generation mix). This has a cost, even if it is perfectly predicted. The perfect forecast is an hourly resolution forecast that perfectly represents the hourly average of the six ten-minute values. It is used as an upper bound on forecast accuracy. The true state-of-the-art forecast will lie between the simplified mesoscale model forecast produced for this project and the perfect forecast produced for this project.

VI. CONCLUSION

This paper discusses the strengths and weaknesses of existing techniques used to develop data for wind integration studies. It also presents a case study using a large wind integration study that employs numerical weather prediction models. The WWSIS is one of the world's largest regional wind integration studies to date. The final dataset covers over 4 million square kilometers, with a spatial resolution of 1 arc-minute and a temporal resolution of ten-minutes. From this dataset over 32,000 sites were selected for further post-processing – each modeled as an individual wind project of 30 MW, resulting in well over 900 GW of synthetic wind energy simulations.

The paper has focused on the process and decisions that formed the basis for the creation of this dataset that will be used for the integration study. The process has been presented with limited consideration of the power systems engineering aspects so that the problems addressed in this paper should be as widely applicable to different integration studies.

VII. ACKNOWLEDGMENT

VIII. REFERENCES

[1] "Utility Wind Integration State of the Art", Utility Wind Integration Group, American Public Power Association, Edison Electric Institute, National Rural Electric Cooperative Association, Reston, VA, USA. May 2006.http://www.uwig.org/opimpactsdocs.html.

7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms

7

[2] "Long-Term Research and Development Needs for Wind Energy For the Time Frame 2000 to 2020", Report to the Executive Committee of the International Energy Agency Implementing Agreement for Co-operation in the Research and Development of Wind Turbine Systems. Oct. 2001.
http://www.ieawind.org/activities_accomplishments.html.

[3] G. C. Larsen and K. S. Hansen, "Database on Wind Characteristics – Analyses of Wind Turbine Design Loads", Risø National Laboratory, Roskilde, Denmark. Feb. 2004.
http://www.ieawind.org/iea_wind_pdf/Appendices/App%20J.pdf.

[4] D. Justus, "International Energy Technology Collaboration and Climate Change Mitigation – Case Study 5: Wind Power Integration into Electricity Systems", Organisation for Economic Co-operation and Development, IEA. May 2005.
www.oecd.org/dataoecd/22/37/34878740.pdf.

[5] "European Wind Integration Study (EWIS) Towards a Successful Integration of Wind Power into European Electricity Grids – Final Report", European Wind Integration Study. Jan 2007.
http://www.wind-integration.eu/downloads/

[6] C. W. Potter, H. A. Gil and J. McCaa, "Wind Power Data for Grid Integration Studies", *Proc. 2007 IEEE Power Engineering Society General Meeting*, Tampa, FL, USA. Paper No. 07GM0808, Jun. 2007.

[7] J. C. Smith, M. R. Milligan, E. A. DeMeo and B. Parsons, "Utility Wind Integration and Operating Impact State of the Art", *IEEE Transactions On Power Systems*, Vol. 22, No. 3, pp. 900-908, Aug. 2007.

[8] T. Boehme, A. R. Wallace and G. P. Harrison, "Applying Time Series to Power Flow Analysis in Networks With High Wind Penetration", *IEEE Transactions On Power Systems*, Vol. 22, No. 3, pp. 951-957, Aug. 2007.

[9] T. Acker, J. Buechler, S. Broome, C. Potter, H. Gil, B. Zavadil, J. Melcher, R. Flood, M. Schmidt, R. La Peter, "Arizona Public Service Wind Integration Cost Impact Study", Northern Arizona University, Sep. 2007
http://www.wind.nau.edu/APSWindIntegrationsStudy.shtml.

[10] "The NCEP/NCAR Reanalysis Project", NOAA/ESRL Physical Sciences Division, Accessed on: Apr. 2008.
http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml

[11] "Northwest Wind Integration Action Plan", Northwest Wind Integration Forum, Northwest Power and Conservation Council, Accessed on: Apr. 2008.
http://www.nwcouncil.org/energy/wind/Default.asp

[12] Skamarock, W. C. et al. "A description of the Advanced Research WRF Version 2", NCAR Tech. Note, NCAR/TN-468+STR, USA, 2005.

[13] G, Grell, J, Dudhia, and D, Stauffer, "A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5)", NCAR Tech. Note, NCAR/TN-398+STR, USA, 1998.

[14] "20% Wind Energy by 2030: Increasing Wind Energy's Contribution to U.S. Electricity Supply", National Renewable Energy Laboratory, U.S. Department of Energy, American Wind Energy Association, Black and Veatch, USA. DOE/GO-102008-2567, Mar 2008.

[15] J. Saintcross et al. "The Effects of Integrating Wind Power on Transmission System Planning, Reliability, and Operations", The New York State Energy Research and Development Authority, Albany, New York, U.S.A. Mar. 2005
http://www.nyserda.org/publications/wind_integration_report.pdf

[16] Enernex Corporation, "Xcel Energy and Minnesota Department of Commence Wind Integration Study – Final Report", Enernex, Knoxville, TN, U.S.A. Sep. 2004.
http://www.nwcouncil.org/energy/Wind/library/EnerexWindIntegrationStudyFinal.pdf

[17] J. Pease, "Challenges With The Integration of Large Scale Wind By A Regional Utility", Bonneville Power Administration. Portland, OR, U.S.A. Jun. 2006.
http://www.bpa.gov/Power/PGC/wind/Challenges_of_large_scale_wind_integration_paper.pdf

[18] Alberta Electric System Operator, "Wind Power Studies", Alberta Electric System Operator. Calgary, Alberta, Canada. Jul. 2006.
http://www.aeso.ca/gridoperations/286.html

[19] R. Piwko, et al, "The Effects of Integrating Wind Power on Transmission System Planning, Reliability and Operations. Report On Phase 2: System Performance Evaluation". Prepared for The New York State Energy Research And Development Authority. GE Energy Consulting, Mar. 2005.
http://www.nyserda.org/publications/wind_integration_report.pdf.

[20] National Center for Environmental Prediction, "Global Climate and Weather Modeling Branch", National Center for Environmental Prediction, Camp Springs, MD, USA. Accessed: Apr 2008.
http://www.emc.ncep.noaa.gov/gmb/

IX. Biographies

Cameron W. Potter graduated from the University of Tasmania, Hobart, Tasmania, Australia with a Ph.D. in Power Systems Engineering and Artificial Intelligence. Cameron also attained his BE with First Class Honours from the University of Tasmania, Hobart, Tasmania, Australia.
Cameron presently works for 3TIER, Inc. as Power Prediction Engineer, which also involves project managing the data creation for wind integration studies. His employment experience also includes working Tutor and Laboratory Supervisor at the University of Tasmania and Technical Assistant at Hydro Tasmania. His special fields of interest include the application of artificial intelligence, data mining and non-deterministic methods to practical renewable energy engineering problems.

Debra Lew graduated with a MS and PhD in Applied Physics from Stanford University, CA, USA, and with BS degrees in Physics and Electrical Engineering from the Massachusetts Institute of Technology, MA, USA.
Debra presently works as a Senior Project Leader in the Systems Integration Team at the National Wind Technology Center at the National Renewable Energy Laboratory (NREL). Her employment experience also includes a dozen years of renewable energy policy, capacity building, and deployment work in developing countries, with a focus on China. She has been at NREL for over ten years and was previously a Senior Energy Advisor at the International Institute for Energy Efficiency in Bangkok, Thailand, and a Postdoctoral Researcher at the Center for Energy and Environmental Studies at Princeton University, NJ, USA. Her fields of interest include high penetrations of wind power in power systems and transmission planning to accommodate wind power.

Jim McCaa graduated from the University of Washington, Seattle, WA, USA with a Ph.D. in Atmospheric Sciences. Jim also has a BA in Physics & Mathematics from Kalamazoo College, Kalamazoo, MI, USA.
Jim presently works for 3TIER, Inc. as Director of High Performance Scientific Computing. His employment experience also includes working previously as Lead Scientist and Research Meteorologist at 3TIER, Atmospheric Model Working Group Liaison at the National Center for Atmospheric Research and Research Associate at the Applied Physics Laboratory, Seattle. His special fields of interest include the physics parameterisations in numerical weather prediction models and management of supercomputing capabilities.

Sam Cheng graduated from Stanford University, Palo Alto, CA, USA with a MS in Electrical Engineering and a BS in Computer Science.
Sam presently works for 3TIER, Inc. as Director of Software Development. His employment experience also includes working previously as a Software Development Engineer at Amazon.com and as a Research Assistant at Stanford University. His special fields of interest include developing scalable software systems for the effective management of large datasets.

Scott Eichelberger graduated from the University of Washington, Seattle, WA, USA with a Ph.D. in Atmospheric Sciences. Scott also has a MS in Atmospheric Sciences from the University of Washington, Seattle, WA, USA and a BS in Meteorology & Mathematics from Texas A&M University, College Station, TX, USA.
Scott presently works for 3TIER, Inc. as Director of Resource Assessment. His employment experience also includes working previously as a Research Scientist at 3TIER and as a Research Assistant at the University of Washington. His special fields of interest include climate dynamics and the influence of climate change on the large-scale circulation of the atmosphere.

Eric Grimit graduated from the University of Washington, Seattle, WA, USA with a Ph.D. in Atmospheric Sciences. Eric also has a MS in Atmospheric Sciences from the University of Washington, Seattle, WA, USA and a BS, *summa cum laude*, in Meteorology from University of Utah, Salt Lake City, UT, USA.
Eric presently works for 3TIER, Inc. as Director of Forecasting. His employment experience also includes working previously as a Numerical Weather Prediction Scientist at 3TIER and as a Research Scientist and Research Assistant at the University of Washington. His special fields of interest include mesoscale numerical weather prediction models, ensemble methods and post-processing.